

## WHAT IS CLAIMED:

1. A method for managed object replication and delivery, comprising:  
directing a request by a client for an object to an edge server in a network;  
if the edge server has the requested object, serving the requested object to the client;  
otherwise, redirecting the client request to a server that has the requested object and serving the requested object to the client;  
if the requested object is popular, replicating the requested object to the edge server.
2. The method of claim 1, wherein redirecting the client request to a server comprises redirecting the client request to a parent server in the network that has the requested object and serving the requested object to the client from the parent server.
3. The method of claim 1, wherein redirecting the client request to a server comprises redirecting the client request to a parent server in the network that does not have the requested object, recursively redirecting the request until a parent server in the network having the requested object is reached and serving the requested object to the client from the parent server.
4. The method of claim 1, wherein redirecting the client request to a server comprises redirecting the client request to an origin server if the requested object is not available at a parent server in the network and serving the requested object to the client from the origin server.

5. The method of claim 1, wherein directing a request by a client for an object to an edge server comprises directing a request by a client for an object to a best or optimal edge server.

6. The method of claim 5, wherein a best or optimal edge server comprises an edge server selected using at least one of a determination based on a best repeater selector, the likelihood of a copy of the requested object being available at the edge server, and the bandwidth between the edge server and the client.

7. The method of claim 1, wherein replicating the requested object to the edge server comprises replicating the requested object to the edge server from a parent server.

8. The method of claim 1, further comprising:

if the requested object is popular and the requested object is unavailable on parent servers in the network, replicating the requested object to a parent server in the network from an origin server.

9. The method of claim 1, wherein replicating the requested object to the edge server comprises, if the requested object is unavailable on parent servers in the network, replicating the requested object to the edge server from an origin server.

10. The method of claim 1, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

11. The method of claim 1, further comprising:

if an object on the edge server is no longer popular, deleting the object from the edge server.

12. The method of claim 1, further comprising:

if an object on the parent server is no longer popular and the object is available on an origin server, deleting the object from the parent server.

13. The method of claim 1, wherein replicating the requested object comprises replicating the requested object in accordance with a dynamic replication threshold.

14. The method of claim 1, wherein replicating the requested object comprises:

replicating the requested object when a popularity of the requested object is greater than a threshold popularity and there is enough storage to replicate the requested object;

otherwise, if there is not enough storage to replicate the requested object,

i) comparing the popularity of the requested object against a popularity of a least popular object in the storage,

ii) if the popularity of the requested object exceeds the popularity of the least popular object in the storage, deleting the least popular object from the storage,

iii) repeating i) and ii) until enough storage is available for the requested object or the popularity of the requested object is less than the popularity of the least popular object in the storage, and

iv) replicating the requested object if there is enough storage.

15. The method of claim 1, wherein serving the requested object is performed separately from replicating the requested object.

16. A method for managed object replication and delivery, comprising:  
directing a request by a client for an object to an optimal edge server in a network;  
if the edge server has the requested object, serving the requested object to the client;

otherwise, redirecting the client request to a parent server in the network that has the requested object and serving the requested object to the client from the parent server;

if the requested object is popular, replicating the requested object to the edge server from the parent server.

17. The method of claim 16, further comprising:

if an object on the edge server is no longer popular, deleting the object from the edge server.

18. The method of claim 16, further comprising:

if an object on the parent server is no longer popular and the object is available on an origin server, deleting the object from the parent server.

19. The method of claim 16, wherein replicating the requested object comprises replicating the requested object in accordance with a dynamic replication threshold.

20. The method of claim 16, wherein replicating the requested object comprises:  
replicating the requested object when a popularity of the requested object is greater than a threshold popularity and there is enough storage to replicate the requested object;  
otherwise, if there is not enough storage to replicate the requested object,  
i) comparing the popularity of the requested object against a popularity of a least popular object in the storage,  
ii) if the popularity of the requested object exceeds the popularity of the least popular object in the storage, deleting the least popular object from the storage,  
iii) repeating i) and ii) until enough storage is available for the requested object or the popularity of the requested object is less than the popularity of the least popular object in the storage, and  
iv) replicating the requested object if there is enough storage.

21. The method of claim 16, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

22. The method of claim 16, wherein serving the requested object is performed separately from replicating the requested object.

23. A computer program product including computer program code to cause a processor to perform a method for managed object replication and delivery, the method comprising:

directing a request by a client for an object to an edge server in a network;  
if the edge server has the requested object, serving the requested object to the client;

otherwise, redirecting the client request to a server that has the requested object and serving the requested object to the client;

if the requested object is popular, replicating the requested object to the edge server.

24. The computer program product of claim 23, wherein redirecting the client request to a server comprises redirecting the client request to a parent server in the network that has the requested object and serving the requested object to the client from the parent server.

25. The computer program product of claim 23, wherein redirecting the client request to a server comprises redirecting the client request to a parent server in the network that does not have the requested object, recursively redirecting the request until a parent server in the network having the requested object is reached and serving the requested object to the client from the parent server.

26. The computer program product of claim 23, wherein redirecting the client request to a server comprises redirecting the client request to an origin server if the requested object is not available at a parent server in the network and serving the requested object to the client from the origin server.

27. The computer program product of claim 23, wherein directing a request by a client for an object to an edge server comprises directing a request by a client for an object to a best or optimal edge server.

28. The computer program product of claim 27, wherein a best or optimal edge server comprises an edge server selected using at least one of a determination based on a best repeater selector, the likelihood of a copy of the requested object(s) being available at the edge server, and the bandwidth between the edge server and the client.

29. The computer program product of claim 23, wherein replicating the requested object to the edge server comprises replicating the requested object to the edge server from a parent server.

30. The computer program product of claim 23, the method further comprising:  
if the requested object is popular and the requested object is unavailable on parent servers in the network, replicating the requested object to a parent server in the network from an origin server.

31. The computer program product of claim 23, wherein replicating the requested object to the edge server comprises, if the requested object is unavailable on parent servers in the network, replicating the requested object to the edge server from an origin server.

32. The computer program product of claim 23, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

33. The computer program product of claim 23, the method further comprising:  
if an object on the edge server is no longer popular, deleting the object from the edge server.

34. The computer program product of claim 23, the method further comprising:  
if an object on the parent server is no longer popular and the object is available on an origin server, deleting the object from the parent server.

35. The computer program product of claim 23, wherein replicating the requested object comprises replicating the requested object in accordance with a dynamic replication threshold.

36. The computer program product of claim 23, wherein replicating the requested object comprises:

replicating the requested object when a popularity of the requested object is greater than a threshold popularity and there is enough storage to replicate the requested object;

otherwise, if there is not enough storage to replicate the requested object,

i) comparing the popularity of the requested object against a popularity of a least popular object in the storage,

ii) if the popularity of the requested object exceeds the popularity of the least popular object in the storage, deleting the least popular object from the storage,

iii) repeating i) and ii) until enough storage is available for the requested object or the popularity of the requested object is less than the popularity of the least popular object in the storage, and

iv) replicating the requested object if there is enough storage.

37. The computer program product of claim 23, wherein serving the requested object is performed separately from replicating the requested object.



38. A computer program product including computer program code to cause a processor to perform a method for managed object replication and delivery, the method comprising:

directing a request by a client for an object to an optimal edge server in a network;  
if the edge server has the requested object, serving the requested object to the client;

otherwise, redirecting the client request to a parent server in the network that has the requested object and serving the requested object to the client from the parent server;

if the requested object is popular, replicating the requested object to the edge server from the parent server.

39. The computer program product of claim 38, the method further comprising:

if an object on the edge server is no longer popular, deleting the object from the edge server.

40. The computer program product of claim 38, the method further comprising:

if an object on the parent server is no longer popular and the object is available on an origin server, deleting the object from the parent server.

41. The computer program product of claim 38, wherein replicating the requested object comprises replicating the requested object in accordance with a dynamic replication threshold.

42. The computer program product of claim 38, wherein replicating the requested object comprises:

replicating the requested object when a popularity of the requested object is greater than a threshold popularity and there is enough storage to replicate the requested object;

otherwise, if there is not enough storage to replicate the requested object,

i) comparing the popularity of the requested object against a popularity of a least popular object in the storage,

ii) if the popularity of the requested object exceeds the popularity of the least popular object in the storage, deleting the least popular object from the storage,

iii) repeating i) and ii) until enough storage is available for the requested object or the popularity of the requested object is less than the popularity of the least popular object in the storage, and

iv) replicating the requested object if there is enough storage.

43. The computer program product of claim 38, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

44. The computer program product of claim 38, wherein serving the requested object is performed separately from replicating the requested object.

45. A system for managed object replication and delivery, comprising:

a plurality of edge servers in a network; and

a plurality of parent servers in the network,

wherein at least one of the plurality of edge servers and the plurality of parent servers:

direct a request by a client for an object to an edge server in the network,

if the edge server has the requested object, serve the requested object to the client,

otherwise, redirect the client request to a server that has the requested object and serve the requested object to the client,

if the requested object is popular, replicate the requested object to the edge server.

46. The system of claim 45, wherein redirect the client request to a server comprises redirect the client request to a parent server in the network that has the requested object and serve the requested object to the client from the parent server.

47. The system of claim 45, wherein redirect the client request to a server comprises redirect the client request to a parent server in the network that does not have the requested object, recursively redirect the request until a parent server in the network having the requested object is reached and serve the requested object to the client from the parent server.

48. The system of claim 45, wherein redirect the client request to a server comprises redirect the client request to an origin server if the requested object is not available at a parent server in the network and serve the requested object to the client from the origin server.

49. The system of claim 45, wherein direct a request by a client for an object to an edge server comprises direct a request by a client for an object to a best or optimal edge server.

50. The system of claim 49, wherein a best or optimal edge server comprises an edge server selected using at least one of a determination based on a best repeater selector,

the likelihood of a copy of the requested object(s) being available at the edge server, and the bandwidth between the edge server and the client.

51. The system of claim 45, wherein replicate the requested object to the edge server comprises replicate the requested object to the edge server from a parent server.

52. The system of claim 45, wherein at least one of the plurality of edge servers and the plurality of parent servers further replicate the requested object from an origin server if the requested object is popular and the requested object is unavailable on parent servers in the network.

53. The system of claim 45, wherein replicate the requested object to the edge server comprises, if the requested object is unavailable on parent servers in the network, replicate the requested object to the edge server from an origin server.

54. The system of claim 45, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

55. The system of claim 45, wherein at least one of the plurality of edge servers and the plurality of parent servers further delete an object if the object is no longer popular.

56. The system of claim 45, wherein replicate the requested object comprises replicate the requested object in accordance with a dynamic replication threshold.

57. The system of claim 45, wherein replicate the requested object comprises:  
replicate the requested object when a popularity of the requested object is greater than a threshold popularity and there is enough storage to replicate the requested object;  
otherwise, if there is not enough storage to replicate the requested object,  
i) compare the popularity of the requested object against a popularity of a least popular object in the storage,  
ii) if the popularity of the requested object exceeds the popularity of the least popular object in the storage, delete the least popular object from the storage,  
iii) repeat i) and ii) until enough storage is available for the requested object or the popularity of the requested object is less than the popularity of the least popular object in the storage, and  
iv) replicate the requested object if there is enough storage.

58. The system of claim 45, wherein serve the requested object is performed separately from replicate the requested object.

59. A system for managed object replication and delivery, comprising:  
a plurality of edge servers in a network; and  
a plurality of parent servers in the network,  
wherein at least one of the plurality of edge servers and the plurality of parent servers:  
direct a request by a client for an object to an optimal edge server in the network;  
if the edge server has the requested object, serve the requested object to the client;  
otherwise, redirect the client request to a parent server in the network that has the requested object and serve the requested object to the client from the parent server;

if the requested object is popular, replicate the requested object to the edge server from the parent server.

60. The system of claim 59, wherein at least one of the plurality of edge servers and the plurality of parent servers further delete an object if the object is no longer popular.

61. The system of claim 59, wherein replicate the requested object comprises replicate the requested object in accordance with a dynamic replication threshold.

62. The system of claim 59, wherein replicate the requested object comprises:  
replicate the requested object when a popularity of the requested object is greater than a threshold popularity and there is enough storage to replicate the requested object;  
otherwise, if there is not enough storage to replicate the requested object,

i) compare the popularity of the requested object against a popularity of a least popular object in the storage,

ii) if the popularity of the requested object exceeds the popularity of the least popular object in the storage, delete the least popular object from the storage,

iii) repeat i) and ii) until enough storage is available for the requested object or the popularity of the requested object is less than the popularity of the least popular object in the storage, and

iv) replicate the requested object if there is enough storage.

63. The system of claim 59, wherein whether the requested object is popular is determined using at least a request rate for the requested object.

64. The system of claim 59, wherein serve the requested object is performed separately from replicate the requested object.

65. The system of claim 59, wherein at least one of the plurality of edge servers and the plurality of parent servers further replicate the requested object from an origin server if the requested object is popular and the requested object is unavailable on parent servers in the network.